DOCUMENT RESUME

ED 410 289	TM 027 129
AUTHOR	Nevitt, Jonathan; Tam, Hak P.
TITLE	A Comparison of Robust and Nonparametric Estimators under the Simple Linear Regression Model.
PUB DATE	Mar 97
NOTE	32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Analysis; Computer Simulation; *Error of Measurement; *Estimation (Mathematics); *Nonparametric Statistics; *Regression (Statistics); *Robustness (Statistics); Sample Size
IDENTIFIERS	*Linear Models; Outliers

ABSTRACT

This study investigates parameter estimation under the simple linear regression model for situations in which the underlying assumptions of ordinary least squares estimation are untenable. Classical nonparametric estimation methods are directly compared against some robust estimation methods for conditions in which varying degrees of outliers are present in the observed data. In addition, estimator performance is considered under conditions in which the normality assumption regarding error distributions is violated. The study addresses the problem through computer simulation methods. The study design includes 3 sample sizes (n=10, 30, 50) crossed with 5 types of error distributions (unit normal, 10% contaminated normal, 30% contaminated normal, lognormal, t-5df). Variance, bias, mean squared error, and relative mean squared error are used to evaluate estimator performance. Recommendations to applied researchers and direction for further study are considered. (Contains 4 tables, 4 figures, and 20 references.) (Author/SLD)

A Comparison of Robust and Nonparametric Estimators Under the Simple Linear Regression Model

Jonathan Nevitt
University of Maryland at College park
Hak P. Tam
National Taiwan Normal University

This paper is prepared for the Annual Meeting of the American Educational Research Association in Chicago, IL

A Comparison of Robust and Nonparametric Estimators Under the Simple Linear Regression Model

Jonathan Nevitt
University of Maryland at College Park
Hak P. Tam
National Taiwan Normal University

Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, March, 1997

The present study investigates parameter estimation under the simple linear regression model for situations in which the underlying assumptions of ordinary least squares (OLS) estimation are untenable. Classical nonparametric estimation methods are directly compared against some robust estimation methods for conditions in which varying degrees of outliers are present in the observed data. Additionally, estimator performance is considered under conditions in which the normality assumption regarding error distributions is violated. The study addresses the problem via computer simulation methods. The study design includes 3 sample sizes (n = 10, 30, 50) crossed with 5 types of error distributions (unit normal, 10% contaminated normal, 30% contaminated normal, lognormal, t-5df). Variance, bias, mean squared error, and relative mean squared error are used to evaluate estimator performance. Recommendations to applied researchers and direction for further study are considered.

Introduction

Applied statistics in the social sciences has focused heavily on modeling data via a linear model (Pedhazur, 1982). Under this framework, a model is posited in which it is assumed that a linear combination of predictors is useful in explaining or predicting some random outcome variable of interest. The most basic form of this model, simple linear regression, is the situation in which a single predictor is included in the explanatory model.

The simple linear regression model, in terms of the observed data, may be expressed by the equation: $Y_i = \alpha + \beta X_i + e_i$, where Y_i is the score for the response measure for the i^{th} individual; X_i is the value of the explanatory variable for the i^{th} individual; α is the Y-intercept, the mean of the population when the value of X is zero; β is the regression coefficient in the population, the slope of the line; e_i is a random disturbance, or error, for individual i. Under this model, it is posited that the score for an individual is partitioned into a structural component ($\alpha + \beta X_i$) which is common to all subjects at the same level of X and a random component (e_i) which is unique to each individual.

In the simple linear regression model, the population parameters α and β are unknown quantities which are estimated from the sample data. The most widely employed method for estimating these parameters is the method of ordinary least squares (OLS). Under the OLS method , the estimates of α and β are chosen to minimize the sum of the squared errors of prediction. OLS regression yields estimates for the parameters that have the desirable property of being minimum variance unbiased estimators (Pedhazur, 1982).

Ordinary least squares estimation places certain restrictive assumptions on the random component in the model, the errors of prediction. OLS estimation assumes, among others, that the errors of prediction are normally distributed, with a common error variance at all levels of X [$e \sim N(0,\sigma^2)$]. These assumptions are frequently untenable in practice. Violations of these assumptions are manifested by the presence of outliers in the observed data. Thus data containing outlying values may reflect nonnormal error distributions with heavy tails or normal error distributions containing observations atypical of the usual normal distribution with larger variance than the assumed σ^2 (Draper & Smith, 1981; Hamilton, 1992). It is well demonstrated that estimates using OLS regression are heavily influenced by outliers in the sample data (Birkes & Dodge, 1993).

It is well recognized that in the presence of normally distributed errors and homoscedasticity, OLS estimation is the method of choice. It is also well documented that estimates using OLS are sensitive to even one outlier in the sample data (see e.g., Rousseeuw & Leroy, 1987). For situations in which the underlying assumptions of OLS estimation are not tenable, the choice of method for parameter estimation is not clearly defined. Thus, the choice of estimation method under non-ideal conditions has been a long standing problem for methodological researchers. The history of this problem is lengthy with many alternative estimation methods having been proposed and investigated (Birkes & Dodge, 1993; Dietz, 1987; Iman & Conover, 1979; Tam, 1996; Theil, 1950; Yale & Forsythe, 1976).

Alternatives to OLS regression may be regarded as falling into broad classes based upon the approach to the problem of parameter estimation and the assumptions placed upon the model. Robust regression is a general term that encompasses a wide array of estimation methods. In general, robust estimation methods are considered to perform reasonably well if the errors of prediction have a distribution that is not necessarily normal but "close" to normal (Birkes & Dodge, 1993). Thus these methods have been developed for situations in which symmetric error distributions have heavy tails due to outliers in the observed data (Hamilton, 1992). A common element to these methods is the definition of a loss function on the residuals and then the minimization of this function for parameter estimation (Draper & Smith, 1981). Examples of this type of robust estimation are Huber M-estimation, the method of Least Absolute Deviations (LAD), and the method of Least Median of Squares.

Other forms of robust regression involve iterative modification of the sample data based upon the residuals from an initial OLS estimation. Examples of this type of robust estimation are Winsorized Least Squares (Yale & Forsythe, 1976) and regression using data trimming methods (Hamilton, 1992). These methods maintain the assumptions of OLS estimation but attempt to smooth the data to resolve the influence of outliers on the parameter estimates.

In distinction to robust regression methods, classical nonparametric approaches to the linear regression problem typically employ parameter estimation methods that are regarded as distribution free. Many nonparametric procedures are based on using the ranks of the observed data rather than the observed data themselves. Nonparametric regression methods are considered to perform well without regard to the nature of the distribution of errors. Since nonparametric regression procedures are developed without relying on the assumption of normality of error distributions, the only presupposition behind such procedures is that the errors of prediction are independently and identically distributed (i.i.d.) (Dietz, 1989). This is a considerably weaker assumption as compared to the assumptions underlying OLS regression and even to those of robust regression procedures.

Several classical nonparametric approaches to the linear regression problem are reviewed by Tam (1996). Some examples of nonparametric methods considered by Tam (1996) are the method of median of pairwise slopes as proposed by Theil (Theil, 1950), a weighted median of pairwise slopes (Jaeckel, 1972), and the rank transformation procedure known as monotonic regression (Iman & Conover, 1979). Additionally, Tam (1996) reviews two important simulation studies by Hussain and Sprent (1983) and by Dietz (1987).

Hussain and Sprent (1983) published a simulation study in which the behavior of several non-parametric estimators was investigated. They compared (among others) the least squares linear regression estimator against the Theil pairwise median and weighted pairwise median estimators in a study using 100 replications per condition. Hussain and Sprent characterized the data modeled in their study as typical data patterns that might result from contamination due to outliers. Contaminated data sets were generated using a mixture model in which each error term is either a random observation from a $N(0,1)$ distribution or an observation from a $N(0,k^2)$ distribution with $k > 1$.

The investigators present results from simulated data sets with the probability, p, of drawing data from the N(0,1) distribution fixed between 0.85 and 0.95. Sample sizes of 10 and 30 are presented for the situation in which there are no outliers ($p = 1.0$) and for the condition in which the data contain approximately 10% outliers ($k = 9$; $p = 0.85$ for $n = 10$, $p = 0.90$ for $n = 30$). X_i values in the Hussain and Sprent study follow an equally spaced, sequential additive series ($X_i = 1, 2, ..., n$). Observed outcome values are generated by the model: $Y_i = 2 + X_i + e_i$, in which e_i is a random deviate drawn from the appropriate normal distribution.

Results from Hussain and Sprent (1983) indicate that Theil's method was appreciably better than OLS in the presence of outliers, especially for small sample sizes. Such results pertain especially to the estimation of the Y-intercept term in the linear regression model. Furthermore, their results showed no real advantage of the weighted median estimator as compared to the Theil estimator under their simulated data conditions.

In addition to the work of Hussain and Sprent, findings in Dietz (1987) have contributed substantially to the field of classical non-parametric regression. Dietz estimated and compared the mean squared errors (MSE's) of the Theil slope and several weighted median slope estimators under a variety of simulated data conditions. Additionally, Dietz examined several nonparametric estimators of Y-intercept. Dietz simulated data according to two sample sizes (20 and 40), three X designs to generate X values, and nine error distributions (i.e. standard normal, 6 contaminated normal distributions with various degrees of flatness, heavy-tailed t distribution with 3 degrees of freedom, and the asymmetric lognormal distribution). Dietz generated 500 data replications per condition.

Findings in Dietz (1987) demonstrated that for normal error distributions, the OLS slope estimator yielded the lowest MSE, while for nonnormal errors the OLS slope estimator had the largest MSE. The weighted median slope estimators showed strong performance under the moderately contaminated data conditions while the Theil unweighted median slope estimator yielded the lowest MSE under the

heavily contaminated data conditions. Dietz also reported that the Y-intercept estimator as proposed by Theil (1950) yielded large MSE's and should be avoided in practice.

Alternatives to OLS regression continue to intrigue applied statisticians and methodological researchers. As a continuation of previous research, our present study explores the behavior of nonparametric approaches to simple linear regression under various situations with respect to contaminated data. This study provides an extension to previous research in some important areas. Primarily, as noted by Tam (1996), very little research exists in which classical nonparametric alternatives to linear regression are directly compared against robust regression methods. Our study serves to address this void in the literature. Additionally, comparisons of alternative regression methods are often presented only within the framework of statistical theory or by examining estimator performance on exemplary data sets (e.g., Birkes & Dodge, 1993). The present study serves to address the issue of comparing alternatives to OLS regression within the framework of a simulation study.

Methods

All programming for the simulation study was developed using GAUSS (Aptech Systems, 1994). In the present study, 3 levels of sample size ($n = 10, 30, 50$) were crossed with 5 types of error distribution (unit normal, unit normal - 10% Y outliers, unit normal - 30% Y outliers, lognormal, t-5df). For each of the 15 cells in the study, 1000 simulated bivariate data sets were generated. Algorithms for drawing random deviates from contaminated unit normal, lognormal, and t-5df distributions are found in Evans, Hastings, and Peacock (1993).

Data generation methods are conformable to those of Hussain and Sprent (1983). Vectors of random error variates were drawn from the appropriate error distribution. Error vectors for the contaminated normal distributions were mixtures of deviates drawn from a unit normal distribution and from a normal $N(0,k^2)$ distribution with $k = 9$.

It has been demonstrated that drawing deviates from this larger variance normal distribution will result in some (potentially) large Y outliers (Hussain & Sprent, 1983).

Simulated bivariate data sets consisted of (X,Y) vectors. The vector of X values was generated to follow an equally spaced, sequential additive series ($X_i = 1, 2, ..., n$). The Y vector was generated by the model: $Y_i = 2 + X_i + e_i$, in which e_i is a random deviate drawn from the appropriate error distribution. Thus, the population parameters underlying the model are $\alpha = 2$ and $\beta = 1$ for Y-intercept and slope respectively.

For each simulated data set, all estimators of slope and intercept were computed. The robust regression estimators considered in this study are LAD, 10% and 20% Winsorized least squares, and 10% trimmed least squares. Algorithms for computing the LAD estimator are found in Birkes and Dodge (1993). Winsorization methods are those developed by Yale and Forsythe (1976). Using the terminology of Yale and Forsythe, Winsorization in the present study is 5 degree using the RES method for parameter estimation. The trimmed least squares estimator is computationally similar to a trimmed mean (Hamilton, 1992). Estimates for 10% trimmed least squares were computed by deleting cases corresponding to the 10% largest positive and the 10% largest negative residuals under an initial OLS estimation. After case deletion, OLS estimation was performed on the remaining simulated data to compute the trimmed least squares estimator.

The classical nonparametric estimators included in this project are monotonic regression (Iman & Conover, 1979), the Theil median estimator (Theil, 1950), and a weighted median estimator (Birkes & Dodge, 1993). Several additional nonparametric intercept estimators investigated by Dietz (1987) were also investigated in the present study.

Summary measures for each estimator were obtained for the set of 1000 replications in each of the 15 cells in the study. Summary measures of minima and maxima, mean, and median were collected. To measure the quality of parameter estimation, estimator variance, bias, mean squared error (MSE) and relative mean squared error (RMSE) were computed for the estimators under each condition. MSE is well recognized as a useful measure of the quality of parameter estimation (Stone, 1996). Mean squared error was computed as $MSE = Var(\theta') + bias(\theta')^2$, in which θ' is an estimate of the population parameter θ.

Relative mean squared error has also been used as a measure of the quality of parameter estimation (e.g. see Yale & Forsythe, 1976). RMSE in our study has been modified from the formulation presented by Yale and Forsythe (1976). In the present study, relative mean squared error is computed as $RMSE = (MSE_{OLS} - MSE_{\theta'})/MSE_{OLS}$. We believe this formulation is useful for comparing estimator performance within a given condition, and is interpreted as a proportionate change from baseline, using the OLS estimator MSE within a given data condition as a baseline value. Positive values of RMSE refer to the proportional reduction in the MSE of a given estimator with respect to OLS estimation. Hence, RMSE is interpreted as a relative measure of performance above that of the OLS estimator.

For each simulated data set, all estimators of slope and intercept were computed. The estimators considered in this study and their labeling are:

<u>Slopes</u>
bls = least squares
blad = least absolute deviations
bmon = monotonic regression
bm = Theil median of pairwise slopes
bwm = weighted median of pairwise slopes
bwin10 = 10% Winsorized least squares
bwin20 = 20% Winsorized least squares
btls = 10% trimmed least squares

Intercepts
als = least squares
amon = monotonic regression
am = Theil median of pairwise Y-intercepts (Thiel,1950)
ac = [median(Y) - bm * median(X)] (Conover, 1980)
a1m = median of (Y_i - bm*X_i) (Birkes & Dodge, 1993)
a1wm = median of (Y_i - bwm*X_i)
a2m = median of pairwise averages of (Y_i - bm*X_i) (Dietz, 1987)
a2wm = median of pairwise averages of (Y_i - bwm*X_i)
awin10 = 10% Winsorized least squares
awin20 = 20% Winsorized least squares
atls = 10% trimmed least squares

Results

Effects of sample size

Across sample sizes, estimator variances (and to some lesser degree bias) decrease with increasing sample size. For example, the variances for the OLS slope estimator under the uncontaminated unit normal distribution are 0.011, 0.00043, and 0.000098 for sample sizes n = 10, 30, and 50 respectively. This pattern of decreasing variance and bias holds for all estimators under all error distributions. The pattern seen in the variances is also exhibited in the estimator MSE. This result is reasonable because $MSE(\theta') = var(\theta') + bias(\theta')^2$.

When considering estimator performance, patterns of variances, bias, and MSE are similar across sample sizes. The results for the n = 30 sample size are intermediate to those for the n = 10 and n = 50 sample sizes and are not reported here.

Slope estimator performance

Tables 1 and 2 present summary results for the estimation of population slope under the unit normal, contaminated normal, and nonnormal error distributions for sample sizes n = 10 and n = 50 respectively. For the OLS slope estimator, note the increase in MSE as the degree of contamination in the data increases. OLS slope

estimator MSE values for the lognormal and t-5df error distributions also show increases as compared to the unit normal error distribution.

Under most conditions, the results for monotonic regression in Tables 1 and 2 show small variances for this slope estimator accompanied by large (in absolute value) bias values. For example, in Table 1, the variance for monotonic regression under the uncontaminated unit normal condition is 0.00096 as compared to the variance for the OLS slope estimator of 0.01115. While monotonic regression yields reduced variances, bias values for this slope estimator can be quite large. Bias values in Table 1 for monotonic regression are often several orders of magnitude higher than the corresponding bias values for the other slope estimators. Note that bias values for monotonic regression are not only large in absolute magnitude, but negative. These negative bias values indicate the monotonic regression slope estimator consistently under estimated the population slope value of $\beta = 1.0$.

Under ideal conditions (unit normal error distribution, no contamination), MSE values in Tables 1 and 2 indicate an inflation in MSE for all robust and nonparametric estimators (with the exception of monotonic regression) as compared to OLS. MSE for these slope estimators are larger than for OLS for this condition and thus corresponding RMSE values are negative. LAD and TLS slope estimators exhibit the largest inflation in MSE as compared to OLS with corresponding reductions in relative estimator performance of approximately 64% for the LAD estimator and 47% (n = 10) and 37% (n = 50) for the TLS estimator.

For the 10% data contamination condition, all robust and nonparametric slope estimators (with the exception of monotonic regression) show strong performance gains with 75-84% decreases in MSE as compared to OLS under this moderate level of data contamination. Comparing estimator performance across the two sample sizes, one sees that performance gains are generally lower for the n = 10 sample size with the exception of the TLS slope estimator. The trimmed least squares slope estimator yields an 83.1% reduction in MSE under the n = 10 sample size and a 74.5% reduction in MSE

under the larger sample size condition.

Under the 30% contamination condition, the LAD slope estimator shows superior performance for both the small and large sample sizes. RMSE values in the two tables indicate reductions in MSE of 80.3% and 88.8% for the n = 10 and n = 50 sample sizes respectively. In this extreme contamination condition, the Theil and weighted median estimators also show strong slope estimator performance. For the n = 50 sample size, slope estimator MSE values for the uncontaminated and contaminated data conditions are plotted in Figure 1. Note that MSE values in the figure are scaled by a scaling factor of 0.0001.

For the lognormal error distribution, the nonparametric Theil and weighted median methods exhibit the strongest performance in both the small and large sample sizes. For the n = 10 sample size, Table 1 reports relative reductions in MSE of 71-72% for these nonparametric estimators. For the large sample size, RMSE values in Table 2 show even higher performance gains with relative reductions in MSE of 83-84%. Close to one another, but running a distant second, are the robust LAD and Winsorized least squares estimators with relative reductions in MSE of about 51% for the small sample size and 60% for the large sample size.

Under the t-5df error distribution, the Winsorized least squares estimators and the nonparametric Theil and weighted median estimators yield only small reductions in MSE relative to the OLS MSE under this condition. Table 2 shows reductions in MSE of about 16% for these estimators under the large sample size while for the small sample size, RMSE values in Table 1 show reductions in MSE of only 2-3%. Figure 2 displays the estimator MSE results from the unit normal, lognormal, and t-5df error distributions for the n = 50 sample size. As in Figure 1, MSE's in Figure 2 are scaled by a factor of 0.0001. Note that the MSE's for the N(0,1) condition in Figure 2 represent the same summary measures as the 0% contaminated data in Figure 1.

Y-Intercept estimator performance

Tables 3 and 4 present summary results for the estimation of population Y-intercept under the unit normal, contaminated normal, and nonnormal error distributions for the small and large sample sizes respectively. Similar to the slope estimator, notice (for both the large and small sample sizes) the OLS Y-intercept estimator yields increases in MSE as the contamination in the data increases. Increased MSE values (as compared to the unit normal error distribution) for OLS are also reported for the nonnormal error distributions. For the small sample size, Table 3 reports the largest MSE for the OLS Y-intercept under the 30% data contamination condition with a value of 12.17. Unlike the small sample size, inspection of MSE's for the OLS Y-intercept in Table 4 reveals the largest MSE value falls under the lognormal error distribution with a reported value of 3.10.

Results for the monotonic regression Y-intercept estimator show extremely poor estimator performance under both the large and small sample sizes. Notice in both Tables 3 and 4 the bias values for the Y-intercept for this estimator under all conditions are large and negative. These negative bias values indicate that the monotonic regression Y-intercept estimator consistently underestimates the population value of $\alpha = 2.0$. For the large sample size, and looking across error distributions, MSE values for the monotonic regression Y-intercept estimator are generally larger than the OLS Y-intercept estimator under similar conditions. Thus most RMSE values in Table 4 for monotonic regression are negative, indicative of a loss in estimator performance as compared to OLS.

Similar to the monotonic regression Y-intercept estimator, the Conover Y-intercept (Conover, 1980) did not perform well. For the small sample size, the Conover Y-intercept shows marginal reductions in MSE as compared to the OLS MSE baseline, but these reductions are not evidenced in Table 4 for the n = 50 sample size. For the large sample size, the Conover Y-intercept yields MSE's that are larger than the corresponding OLS MSE's. Thus, RMSE values in Table 4 for the Conover Y-intercept are negative.

Under the uncontaminated, unit normal error distribution, all robust and nonparametric Y-intercept estimators yield an inflation in MSE as compared to OLS. These inflated MSE's are seen for both sample sizes in the two tables. Discounting the monotonic regression and Conover Y-intercept estimators, the LAD and TLS estimators exhibit the most dramatic loss in estimator performance.

Under the 10% data contamination all nonparametric and robust Y-intercept estimators show strong performance relative to OLS. Discounting the monotonic regression and Conover intercepts, all Y-intercept estimators under both sample sizes yield reductions in MSE of 75-83%. The Y-intercept estimators based on the nonparametric Theil and weighted median tend to have slight advantage over the robust estimators. Also, notice the TLS estimator shows weaker performance in the large sample size condition as compared to the n = 10 sample size cell for this moderately contaminated data condition.

For the 30% contamination, the LAD Y-intercept estimator and the unweighted Theil median pairwise Y-intercept estimator (am) yield the lowest MSE values with the other nonparametric Y-intercepts all very close. These results hold for both the small sample size MSE's in Table 3 and for the n = 50 sample size presented in Table 4.

Under the lognormal error distribution, all estimators of Y-intercept had difficulty in recovering the population value of $\alpha = 2.0$. Note the large bias values for the estimators under this condition, suggesting large discrepancies between the means for the estimators and the population value. The unweighted Theil Y-intercept estimator (am) shows the strongest relative performance under both sample sizes. The nonparametric techniques a1m and a1wm also yield relative strong estimator performance with RMSE values of 0.64 for the n = 50 sample size. For the large sample size, the LAD Y-intercept estimator was also competitive.

For the t-5df error distribution, Tables 3 and 4 report only modest reductions in MSE as compared to the OLS MSE benchmark. Table 3 shows increases in MSE for the LAD estimator as well as for most of the other Y-intercept estimators. For the large sample size, the nonparametric pairwise methods a2m, a2wm demonstrate superior performance with the Winsorized least squares methods also strong. Relative to OLS, these nonparametric methods show a 16.7% decrease in MSE while the Winsorized intercepts exhibit about a 14% reduction in MSE as compared to OLS. The LAD Y-intercept estimator exhibited poor performance with a MSE value slightly larger than that of the OLS Y-intercept estimator.

Discussion

Findings in the present study have substantive implications for educational researchers and research methodologists. The poor performance of OLS estimation under the contaminated data conditions and nonnormal error distributions serves to reaffirm both the importance of assessing underlying assumptions as part of any regression analysis and the need for alternatives to OLS regression. This study has also replicated past findings which have suggested that when the appropriate assumptions are met, OLS regression is the method of choice. Our results have shown, under all sample sizes and for estimation of both population slope and Y-intercept, the OLS estimator yields the lowest mean squared error under ideal conditions.

Findings in the present study have also demonstrated the merits of alternatives to OLS regression under non-ideal conditions. Our results also indicate that estimator performance is dependent upon the nature of the error distribution. Figure 1 shows that under mild (10%) data contamination there is no real preference for one alternative slope estimator over another. When the degree of data contamination is increased to 30% the LAD robust slope estimator moderately outperformed the other slope estimators.

For the case of nonnormal error distributions, our results demonstrate that the symmetry of the error distribution substantially impacts estimator performance. Figure 2 illustrates that when the error distribution is nonnormal and symmetric (t-5df errors) the robust LAD estimator, which demonstrated strong performance under the contaminated normal conditions, is not a desirable choice. Under this

condition, the Winsorized least squares and nonparametric methods employing medians of pairwise slopes (Theil and modified Theil) exhibited superior performance. Figure 2 also demonstrates that when the error distribution is skewed, the nonparametric Theil methods yield very strong performance.

The monotonic regression and the trimmed least squares methods investigated in this study were generally not competitive. The poor results obtained for monotonic regression are not entirely unexpected. In their proposal of this alternative method of regression, Iman and Conover (Iman & Conover, 1979) caution the use of this method under situations in which there are outliers in the observed data. They recommend this method only for situations in which observed data exhibits a monotonically increasing or decreasing trend - curvilinear data. Additionally other investigators have found the rank transformation procedure to be problematic (Sawilowsky, Blair & Higgins, 1989; McKean & Vidmar, 1994). Our results have served to substantiate these findings with empirical evidence of the unacceptability of rank transformation in the form of monotonic regression with respect to bias and RMSE. Large bias values in the summary tables reflect monotonic regression's inability to recover the true population values under our data conditions.

The results for monotonic regression in this study also provide valuable insight into the use of the MSE as a sole indicator of the quality of parameter estimation. For us, a useful estimator is one in which both variance and bias are minimized. Figure 2 shows monotonic regression as having very low MSE under the N(0,1) and t-5df error distributions. The small values for monotonic regression in this figure can be misleading with respect to choice of estimator. Table 2 reports bias values for monotonic regression that are approximately 10 times larger than the bias values for the other slope estimators under each condition. We present Figure 3 which charts bias values for the various estimators under the unit normal, t-5df, and lognormal error distributions for the $n = 50$ sample size. When considering bias as a measure of the quality of parameter estimation, this figure readily demonstrates that monotonic regression is not an

optimal estimator under the conditions of our study. For clarity of presentation, we also present Figure 4 which presents the same results of Figure 3 with the monotonic regression estimator removed. With respect to assessing the quality of parameter estimation, our recommendation for methodological researchers is to evaluate MSE with the caveat that bias should also be simultaneously considered.

The trimmed least squares estimator was included in the study to address the issue of case deletion, an approach frequently suggested in applied scenarios in which there are outliers in the observed data. For the trimmed least squares estimator, data points corresponding to the 10% largest positive and the 10% largest negative residuals from an initial OLS regression were deleted. Under the contaminated data conditions in this study, the case deletion approach to estimation of population slope did not generate unattractive results, although comparison of the TLS slope estimator in Tables 1 and 2 reveals the performance of this estimator to be sample size dependent. Under the small sample size, the TLS slope estimator performed well under the 10% data contamination, but not under the 30% contamination condition. For the larger sample size, Table 2 reports weaker performance under the moderate contamination condition (with respect to the other slope estimators) but stronger performance under the more extreme 30% data contamination condition. While the performance of the TLS slope estimator was not unreasonable, for both the 10% and 30% contamination conditions, robust and nonparametric methods (discounting monotonic regression) which utilize all the available data outperformed the trimmed least squares method. Additionally, for the conditions in which the distribution of errors was nonnormal, the TLS slope estimator was not competitive. Figure 4 shows very low bias for this estimator, but the variance for this slope estimator tends to be inflated. Thus the MSE's for TLS shown in Figure 2 tend to be higher than some of the other slope estimators. Our results demonstrate that methods which utilize all available data, but are resistant to outlying values, provide more accurate long run estimates of true population values. This conclusion is consistent with previous research in resistant methods of regression (Birkes & Dodge, 1993; Rousseeuw &

Leroy, 1987).

With respect to the estimators investigated in the present study, our results have demonstrated that the nonparametric approaches based on the Theil method of median of pairwise slopes are very strong alternatives to OLS regression. This conclusion holds for the small sample size investigated here as well for the large sample size. This study has demonstrated that this approach provides accurate estimates of true population parameters under both outlier contaminated data conditions and under nonnormal error distributions. While the Theil type methods did not outperform the LAD estimator under the heavily contaminated conditions (30% outliers), these methods were nearly as strong as the LAD regression method under this condition. Under the nonnormal error conditions, no estimator outperformed the Theil methods. Additionally, under the lognormal error distribution, the Theil based regression methods showed superior performance. Thus, the Theil based estimation methods were never the worst, sometimes nearly the best and in some cases the best methods for parameter estimation under the simple linear model.

The Theil based method for parameter estimation has found little attention in social science research and deserves further consideration by applied researchers. These nonparametric methods possess several desirable qualities. First, the present study has demonstrated that the Thiel based regression methods provide strong parameter estimation under a variety of non-ideal conditions. Additionally, there is literature available which provides an extension of this method - using a weighted form of the Theil method - to the multiple predictor situation (Birkes & Dodges, 1993). Included in this literature are hypothesis testing procedures for testing both model adequacy and individual regression coefficients. These tests have been developed within the framework of ordinary Z tests (Birkes & Dodge, 1993). Finally, the modified form of the Theil regression method has been incorporated into at least one of the commonly available applied statistics packages available for researchers. The program Minitab contains the RANK REGRESSION command which performs nonparametric regression estimation based on the weighted Theil

method.

We recommend the following approach to applications in educational research. First, applications should always involve checking for outliers in the observed data and testing the underlying assumptions under OLS estimation. Secondly, researchers may be well advantaged to routinely run both OLS and the Theil based methods when conducting regression based analyses. Should the assumptions of normality and homoscedasticity hold, researchers may adopt and report OLS estimates in their findings. Under applied settings in which the OLS assumptions are not tenable, researchers may feel confidence in the estimates of population values using the nonparametric Theil based method.

The present study only considered estimators under the simple linear regression situation. Future studies might compare the performance of nonparametric Theil based estimators against robust regression estimators under the multiple regression situation. In addition, future studies might be warranted to compare the nonparametric Theil based estimators against robust regression methods such as M-regression (Birkes & Dodge, 1993), iteratively reweighted least squares (Holland & Welsch, 1977), or least median squares regression (Rousseeuw & Leroy, 1987). These robust methods are known to be resistant to more extreme forms of data contamination such as leverage points. Finally, additional research investigating power and Type I error rates using the nonparametric Theil based methods would be useful to more fully characterize the behavior of these methods under hypothesis testing paradigms.

References

Aptech Systems. (1994). *GAUSS System and Graphics Manual.* Aptech Systems, Inc, Maple Valley, WA.

Birkes, D., & Dodge, Y. (1993). *Alternative Methods of Regression.* New York: Wiley.

Conover, W.J. 1980. *Practical Nonparametric Statistics* (2nd edition). New York: Wiley.

Dietz, E. J. (1987). A comparison of robust estimators in simple linear regression. *Communication in Statistics-Simulation, 16,* 1209-1227.

Dietz, E. J. (1989). Teaching regression in a nonparametric statistics course. *The American Statistician, 43,* 35-40.

Draper, N., & Smith, H. (1981). *Applied Regression Analysis* (2nd ed.). New York: Wiley.

Evans, M., Hastings, N., & Peacock, B. (1993). *Statistical Distributions* (2nd ed.). New York: Wiley.

Hamilton, Lawrence C. (1992). *Regression with Graphics, A second Course in Applied Statistics.* Pacific Grove, CA: Brooks/Cole.

Holland, P.H., & Welsch, R.E. (1977). Robust regression using iteratively reweighted least squares. *Communications Statistics: Theory and Methods, 6,* 813-827.

Hussain, S. S., & Sprent, P. (1983). Nonparametric regression. *Journal of the Royal Statistical Society, series A, 146,* 182 - 191.

Iman, R. L., & Conover, W. J. (1979). The use of rank transformation in regression. *Technometrics, 21(4),* 499-509.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *Annals of Mathematical Statistics, 43,* 1449-1458.

McKean, J. W., & Vidmar, T. J. (1994). A comparison of two rank-based methods for the analysis of linear models. *The American Statistician, 48(3),* 220-229.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research* (2nd ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

Sawilowsky, S. S., Blair, R. C., & Higgins, J. J. (1989). An investigation of type I error and power properties of the rank transform procedure in factorial anova. *Journal of Educational Statistics, 14(3),* 255-267.

Stone, C. J. (1996). *A Course in Probability and Statistics.* Belmont, CA: Duxbury.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae, 12,* 85-91.

Tam, H. P. (1996, April). *A review of nonparametric regression techniques.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Yale, C., & Forsythe, A. B. (1976). Winsorized regression. *Technometrics, 18,* 291 - 300.

Table 1. Results for Estimation of Population Slope (β = 1.0) for n = 10 sample size

error distribution = unit normal, 0% contamination

param	Variance	Bias	MSE	RMSE
bls :	0.01115491	0.00707727	0.01120500	0.0
blad :	0.01838679	0.00598824	0.01842265	-0.64414576
bwin10:	0.01223615	0.00756652	0.01229340	-0.09713513
bwin20:	0.01299585	0.00830138	0.01306476	-0.16597602
btls :	0.01646757	0.00737854	0.01652201	-0.47452125
bmon :	0.00096072	-0.04701818	0.00317143	0.71696304
bm :	0.01266696	0.00790564	0.01272946	-0.13605202
bwm :	0.01235103	-0.00126754	0.01235263	-0.10242155

error distribution = unit normal, 10% contamination

param	Variance	Bias	MSE	RMSE
bls :	0.11142026	0.01378250	0.11161021	0.0
blad :	0.02767390	0.00432905	0.02769264	0.75188074
bwin10:	0.02192931	0.00375534	0.02194342	0.80339239
bwin20:	0.02942458	0.00682076	0.02947111	0.73594615
btls :	0.01880606	0.00268830	0.01881329	0.83143757
bmon :	0.02047459	-0.15438788	0.04431021	0.60299146
bm :	0.02066901	0.00651707	0.02071149	0.81443018
bwm :	0.02018951	-0.00604903	0.02022610	0.81877913

error distribution = unit normal, 30% contamination

param	Variance	Bias	MSE	RMSE
bls :	0.31264452	-0.00547711	0.31267452	0.0
blad :	0.06165909	-0.00054303	0.06165939	0.80280009
bwin10:	0.14933177	-0.01329565	0.14950854	0.52183970
bwin20:	0.10990528	-0.00357852	0.10991809	0.64845845
btls :	0.15258114	-0.01516154	0.15281101	0.51127769
bmon :	0.04915750	-0.34893333	0.17091197	0.45338696
bm :	0.06853707	-0.00716128	0.06858835	0.78063978
bwm :	0.09594470	-0.02908675	0.09679074	0.69044252

error distribution = lognormal

param	Variance	Bias	MSE	RMSE
bls :	0.05361053	0.00528236	0.05363843	0.0
blad :	0.02574529	-0.00334989	0.02575651	0.51981235
bwin10:	0.02661642	-0.00448754	0.02663656	0.50340532
bwin20:	0.02639584	0.00045408	0.02639604	0.50788934
btls :	0.03613776	-0.00986773	0.03623513	0.32445574
bmon :	0.01326078	-0.10921212	0.02518806	0.53041014
bm :	0.01489242	-0.00264993	0.01489945	0.72222444
bwm :	0.01521499	-0.01259078	0.01537352	0.71338612

error distribution = t-5df

param	Variance	Bias	MSE	RMSE
bls :	0.01764455	-0.00219277	0.01764936	0.0
blad :	0.02363482	0.00078222	0.02363543	-0.33916683
bwin10:	0.01707596	-0.00241412	0.01708179	0.03215808
bwin20:	0.01734658	-0.00224915	0.01735164	0.01686858
btls :	0.02184069	-0.00112768	0.02184196	-0.23755002
bmon :	0.00321083	-0.07123636	0.00828545	0.53055218
bm :	0.01810661	-0.00002527	0.01810661	-0.02590776
bwm :	0.01704933	-0.01184856	0.01718971	0.02604293

Table 2. Results for Estimation of Population Slope ($\beta = 1.0$) for n = 50 sample size

error distribution = unit normal, 0% contamination

param	Variance	Bias	MSE	RMSE
bls :	0.00009810	0.00027520	0.00009818	0.0
blad :	0.00016128	0.00031704	0.00016138	-0.64382254
bwin10:	0.00010321	0.00022429	0.00010326	-0.05175852
bwin20:	0.00010363	0.00022180	0.00010367	-0.05600182
btls :	0.00013426	0.00001423	0.00013426	-0.36749649
bmon :	0.00000043	-0.00214771	0.00000504	0.94866569
bm :	0.00010445	0.00024160	0.00010451	-0.06451524
bwm :	0.00010419	0.00015729	0.00010421	-0.06148768

error distribution = unit normal, 10% contamination

param	Variance	Bias	MSE	RMSE
bls :	0.00088268	0.00146208	0.00088482	0.0
blad :	0.00017786	0.00016588	0.00017789	0.79895156
bwin10:	0.00015842	0.00036745	0.00015855	0.82081015
bwin20:	0.00019381	0.00049379	0.00019406	0.78068165
btls :	0.00022498	0.00053138	0.00022527	0.74541118
bmon :	0.00010839	-0.01713325	0.00040194	0.54574057
bm :	0.00014566	0.00026384	0.00014573	0.83529971
bwm :	0.00014591	0.00017933	0.00014594	0.83506178

error distribution = unit normal, 30% contamination

param	Variance	Bias	MSE	RMSE
bls :	0.00255733	0.00110911	0.00255856	0.0
blad :	0.00028630	0.00049167	0.00028655	0.88800458
bwin10:	0.00086501	0.00036565	0.00086514	0.66186374
bwin20:	0.00084013	0.00011337	0.00084015	0.67163290
btls :	0.00047307	0.00037883	0.00047321	0.81504716
bmon :	0.00032382	-0.04740485	0.00257104	-0.00487937
bm :	0.00034353	0.00008385	0.00034354	0.86573035
bwm :	0.00034969	0.00000300	0.00034969	0.86332687

error distribution = lognormal

param	Variance	Bias	MSE	RMSE
bls :	0.00041453	-0.00022877	0.00041458	0.0
blad :	0.00015974	0.00023784	0.00015979	0.61456832
bwin10:	0.00016734	-0.00003671	0.00016734	0.59635229
bwin20:	0.00016585	-0.00005373	0.00016586	0.59994409
btls :	0.00025751	0.00016274	0.00025753	0.37881494
bmon :	0.00008034	-0.00906487	0.00016251	0.60800790
bm :	0.00006711	-0.00000821	0.00006711	0.83811642
bwm :	0.00006952	-0.00015675	0.00006954	0.83225400

error distribution = t-5df

param	Variance	Bias	MSE	RMSE
bls :	0.00015653	0.00035306	0.00015665	0.0
blad :	0.00016443	0.00007171	0.00016443	-0.04966430
bwin10:	0.00013112	0.00026626	0.00013119	0.16251913
bwin20:	0.00013395	0.00026166	0.00013402	0.14449768
btls :	0.00015108	-0.00001635	0.00015108	0.03555771
bmon :	0.00000383	-0.00367558	0.00001734	0.88929977
bm :	0.00013153	0.00029653	0.00013162	0.15980954
bwm :	0.00013104	0.00015973	0.00013106	0.16336140

Table 3. Results for Estimation of Population Intercept ($\alpha = 2.0$) for n = 10 sample size

error distribution = unit normal, 0% contamination

param	Variance	Bias	MSE	RMSE
als :	0.46623625	-0.04029911	0.46786027	0.0
alad :	0.69815192	-0.03000075	0.69905197	-0.49414688
awin10:	0.49958621	-0.04058599	0.50123343	-0.07133146
awin20:	0.53987082	-0.04371208	0.54178157	-0.15799866
atls :	0.63496493	-0.04029981	0.63658900	-0.36063915
amon :	0.02906177	-1.74140000	3.06153573	-5.54369673
am :	0.60101301	-0.04286027	0.60285001	-0.28852576
ac :	0.75428972	-0.05273348	0.75707054	-0.61815522
a1m :	0.55398222	-0.03755454	0.55539257	-0.18709068
a2m :	0.51962863	-0.04548496	0.52169752	-0.11507120
a1wm :	0.54440377	0.01445567	0.54461273	-0.16404996
a2wm :	0.50696225	0.00284416	0.50697034	-0.08359348

error distribution = unit normal, 10% contamination

param	Variance	Bias	MSE	RMSE
als :	4.26531434	-0.10433764	4.27620068	0.0
alad :	0.97412322	-0.01615455	0.97438419	0.77213787
awin10:	0.82937738	-0.02296355	0.82990470	0.80592475
awin20:	1.03965787	-0.03238871	1.04070690	0.75662814
atls :	0.69793756	-0.01368506	0.69812484	0.83674180
amon :	0.61935638	-1.15086667	1.94385047	0.54542581
am :	0.79369545	-0.02981975	0.79458466	0.81418443
ac :	1.21925681	-0.06955337	1.22409448	0.71374251
a1m :	0.76404907	-0.02949380	0.76491895	0.82112183
a2m :	0.77021114	-0.03544536	0.77146751	0.81959043
a1wm :	0.75465027	0.03493991	0.75587107	0.82323770
a2wm :	0.75285108	0.03346361	0.75397089	0.82368206

error distribution = unit normal, 30% contamination

param	Variance	Bias	MSE	RMSE
als :	12.16682218	0.06720657	12.17133890	0.0
alad :	2.25787505	0.01654912	2.25814892	0.81446997
awin10:	6.31690620	0.12281904	6.33199072	0.47976219
awin20:	4.06190003	0.04919058	4.06431975	0.66607456
atls :	5.68178363	0.11983283	5.69614354	0.53200354
amon :	1.48701426	-0.08086667	1.49355368	0.87728929
am :	2.28520296	0.04164861	2.28693757	0.81210468
ac :	3.82005669	0.09478596	3.82904107	0.68540511
a1m :	2.52240748	0.07140465	2.52750611	0.79233952
a2m :	2.71451183	0.07882679	2.72072549	0.77646457
a1wm :	3.33977214	0.19174151	3.37653694	0.72258295
a2wm :	3.51903642	0.19300923	3.55628899	0.70781448

error distribution = lognormal

param	Variance	Bias	MSE	RMSE
als :	1.97547147	1.61204928	4.57417434	0.0
alad :	1.00661028	1.17225486	2.38079173	0.47951443
awin10:	1.11938148	1.46240345	3.25800534	0.28773914
awin20:	1.08117248	1.31667353	2.81480166	0.38463175
atls :	1.38701960	1.39707222	3.33883039	0.27006928
amon :	0.40113847	-1.39933333	2.35927225	0.48421899
am :	0.76859984	1.07407962	1.92224688	0.57976091
ac :	1.17645633	1.53385117	3.52915574	0.22846060
a1m :	0.69420592	1.14652836	2.00873319	0.56085338
a2m :	0.74321559	1.31729607	2.47848453	0.45815696
a1wm :	0.72207252	1.20571787	2.17582810	0.52432331
a2wm :	0.76095906	1.37362165	2.64779550	0.42114242

error distribution = t-5df

param	Variance	Bias	MSE	RMSE
als :	0.66246365	0.01522071	0.66269532	0.0
alad :	0.89415823	0.00021390	0.89415828	-0.34927507
awin10:	0.63214117	0.01585090	0.63239242	0.04572674
awin20:	0.64452998	0.01393173	0.64472407	0.02711842
atls :	0.77244571	0.00765923	0.77250437	-0.16570065
amon :	0.09712767	-1.60820000	2.68343491	-3.04927396
am :	0.75091053	0.00304771	0.75091982	-0.13312981
ac :	1.00145486	-0.02732700	1.00220163	-0.51231131
a1m :	0.71263706	0.00291803	0.71264557	-0.07537438
a2m :	0.67747509	0.00604977	0.67751169	-0.02235773
a1wm :	0.67256536	0.06871781	0.67728750	-0.02201944
a2wm :	0.63839635	0.06975511	0.64326212	0.02932449

Table 4. Results for Estimation of Population Intercept ($\alpha = 2.0$) for n = 50 sample size

error distribution = unit normal, 0% contamination

param	Variance	Bias	MSE	RMSE
als :	0.08306252	-0.01121545	0.08318831	0.0
alad :	0.13422339	-0.01715318	0.13451762	-0.61702554
awin10:	0.08685171	-0.01015017	0.08695474	-0.04527592
awin20:	0.08794651	-0.00868562	0.08802195	-0.05810480
atls :	0.10876400	-0.00432918	0.10878275	-0.30766868
amon :	0.00027777	-1.94523347	3.78421102	-44.48969748
am :	0.10683096	-0.01485406	0.10705160	-0.28685873
ac :	0.43071144	-0.00363965	0.43072469	-4.17770702
a1m :	0.09617646	-0.01358010	0.09636088	-0.15834637
a2m :	0.08781186	-0.01081533	0.08792884	-0.05698550
a1wm :	0.09576219	-0.01185850	0.09590282	-0.15284007
a2wm :	0.08751654	-0.00873591	0.08759285	-0.05294667

error distribution = unit normal, 10% contamination

param	Variance	Bias	MSE	RMSE
als :	0.72959431	-0.03815976	0.73105048	0.0
alad :	0.14628716	-0.00046989	0.14628738	0.79989428
awin10:	0.13127624	-0.00755356	0.13133329	0.82034990
awin20:	0.15145369	-0.01096891	0.15157401	0.79266273
atls :	0.17187647	-0.01173481	0.17201417	0.76470275
amon :	0.07048059	-1.56310204	2.51376858	-2.43857044
am :	0.13115023	-0.00677069	0.13119607	0.82053760
ac :	0.97688948	0.00936086	0.97697710	-0.33640170
a1m :	0.12444677	-0.00521617	0.12447398	0.82973272
a2m :	0.12000176	-0.00529341	0.12002979	0.83581191
a1wm :	0.12399224	-0.00303475	0.12400145	0.83037909
a2wm :	0.11985011	-0.00315264	0.11986005	0.83604409

error distribution = unit normal, 30% contamination

param	Variance	Bias	MSE	RMSE
als :	2.22128658	0.00799708	2.22135053	0.0
alad :	0.25062945	-0.01285950	0.25079481	0.88709805
awin10:	1.01203716	0.02938913	1.01290088	0.54401574
awin20:	0.72494948	0.02057523	0.72537282	0.67345414
atls :	0.48621344	0.01482624	0.48643326	0.78101914
amon :	0.21056519	-0.79117633	0.83652517	0.62341596
am :	0.25423256	-0.01322216	0.25440739	0.88547175
ac :	2.36044148	0.01979859	2.36083347	-0.06279195
a1m :	0.28495314	-0.00119235	0.28495456	0.87172013
a2m :	0.30048783	0.00552801	0.30051839	0.86471366
a1wm :	0.28922996	-0.00037240	0.28923010	0.86979538
a2wm :	0.30494324	0.00740885	0.30499813	0.86269698

error distribution = lognormal

param	Variance	Bias	MSE	RMSE
als :	0.38500603	1.64740505	3.09894942	0.0
alad :	0.14386289	1.01840542	1.18101249	0.61889908
awin10:	0.16699356	1.38154934	2.07567214	0.33020135
awin20:	0.15463921	1.28493558	1.80569865	0.41731910
atls :	0.21240544	1.25722778	1.79302712	0.42140807
amon :	0.05224150	-1.76884571	3.18105666	-0.02649519
am :	0.09562958	0.90239609	0.90994829	0.70636878
ac :	0.64892005	1.64056750	3.34038177	-0.07790781
a1m :	0.07920296	1.01310474	1.10558416	0.64323904
a2m :	0.08361419	1.22682737	1.58871960	0.48733607
a1wm :	0.08168383	1.01704798	1.11607041	0.63985523
a2wm :	0.08588540	1.23072013	1.60055744	0.48351611

error distribution = t-5df

param	Variance	Bias	MSE	RMSE
als :	0.12955685	-0.00977434	0.12965239	0.0
alad :	0.13381946	-0.00339975	0.13383102	-0.03222951
awin10:	0.11127929	-0.00778139	0.11133984	0.14124342
awin20:	0.11042154	-0.00831868	0.11049074	0.14779248
atls :	0.12116747	-0.00114135	0.12116877	0.06543357
amon :	0.00249143	-1.90627265	3.63636686	-27.04704888
am :	0.11836755	-0.00596709	0.11840316	0.08676456
ac :	0.56742654	-0.02887466	0.56826029	-3.38295273
a1m :	0.11431813	-0.00951134	0.11440859	0.11757436
a2m :	0.10786731	-0.00879750	0.10794471	0.16742983
a1wm :	0.11455410	-0.00583959	0.11458820	0.11618908
a2wm :	0.10789910	-0.00524933	0.10792666	0.16756907

Figure 1. Mean Squared Errors for Estimation of Population Slope Under Varying Levels of Data Contamination.

N = 50.

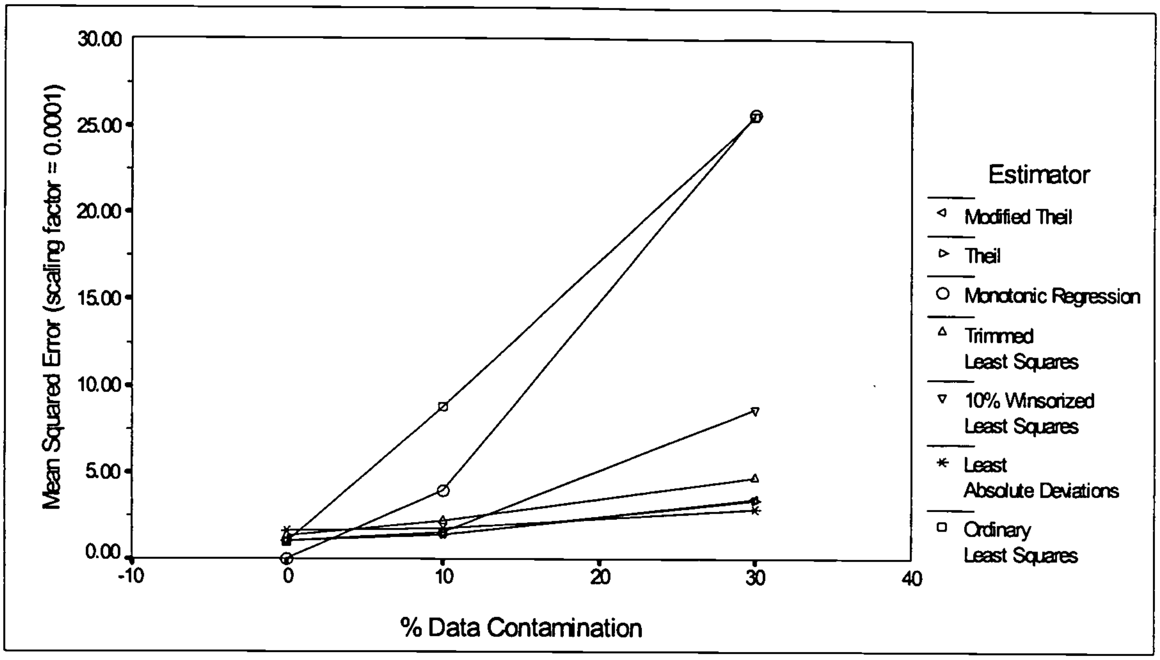


Figure 2. Mean Squared Errors for Estimation of Population Slope for Unit Normal and Nonnormal Error Distributions.

N = 50.

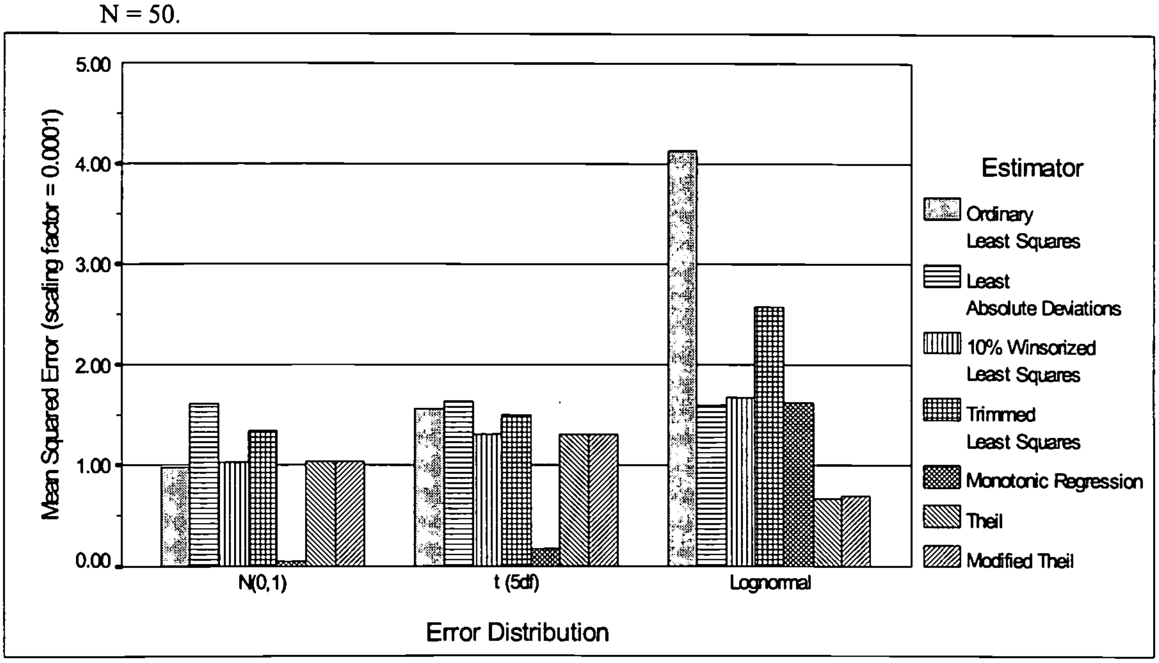


Figure 3. Bias for Estimation of Population Slope for Unit Normal and Nonnormal Error Distributions.

N = 50.

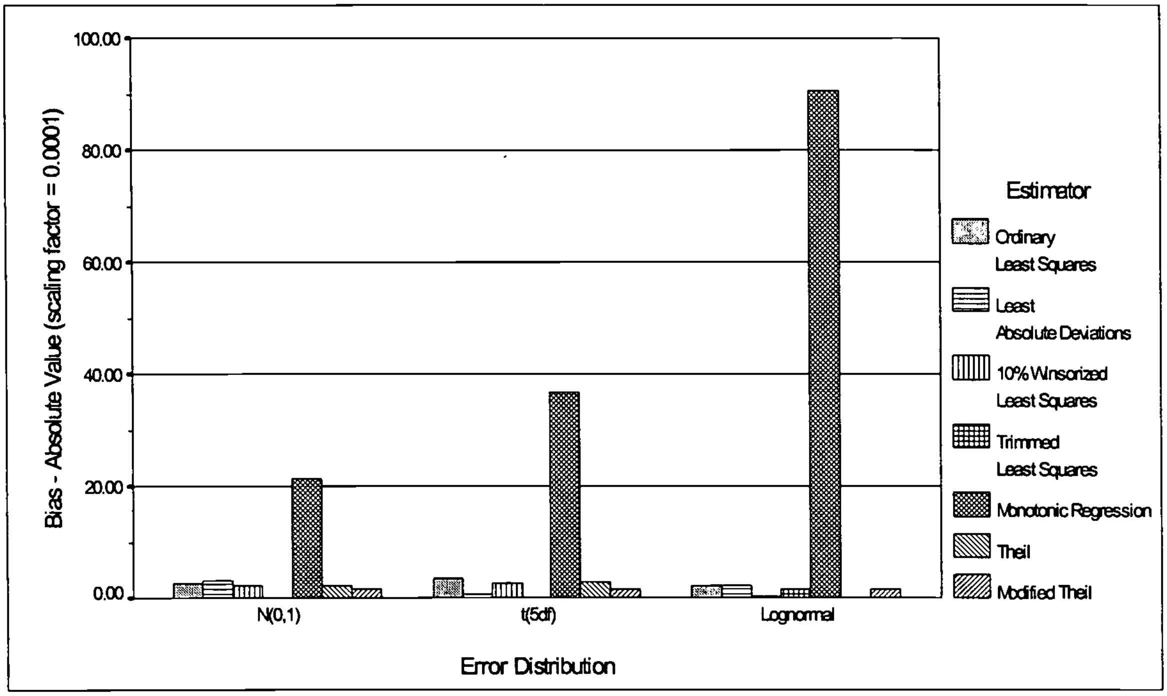
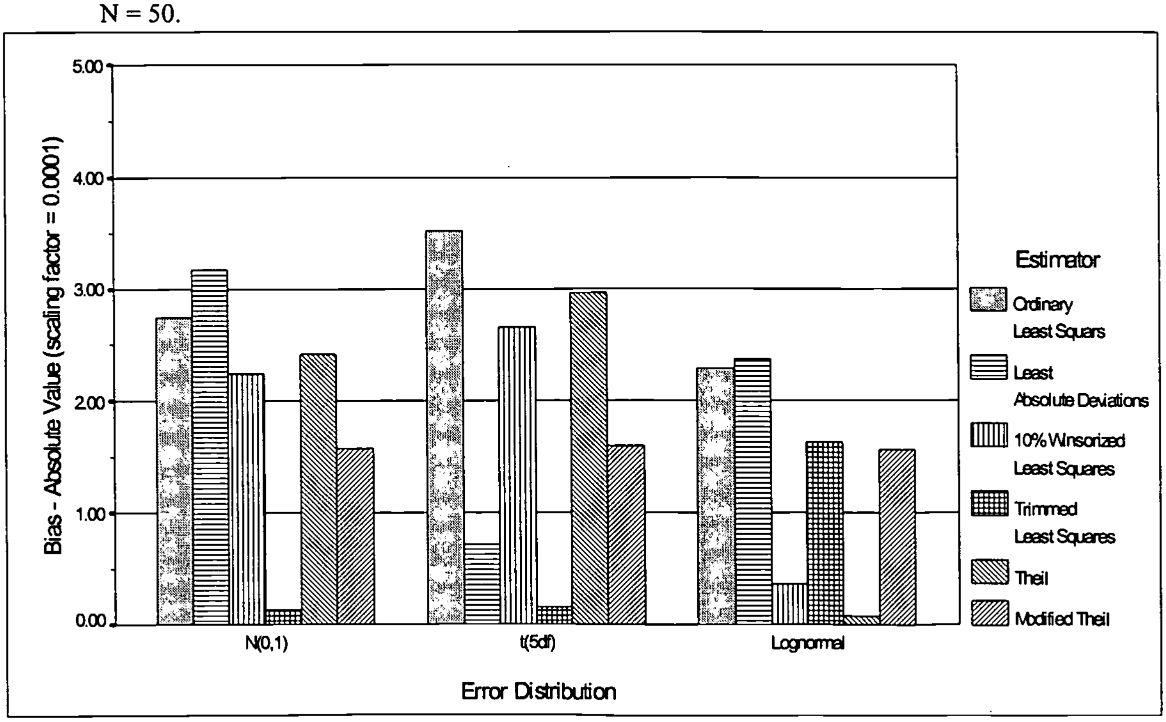


Figure 4. Bias for Estimation of Population Slope for Unit Normal and Nonnormal Error Distributions-Monotonic Regression Slope Estimator Removed.

N = 50.



AREA 1997

TM027129

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A Comparison of Robust & Nonparametric Estimators Under the Simple Linear Regression Model

Author(s): Jonathan Nevitt, Hak P. Tam

Corporate Source:

Publication Date: March 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system. *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] ⬅ Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document ➡ [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: J. Nevitt	Position: Graduate Student
Printed Name: Jonathan Nevitt	Organization: University of MD at College Park
Address: 7656 Stony Creek Lane Ellicott City, MD 21043	Telephone Number: (410) 796-3720
	Date: 4/30/97

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 21, 1997

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (523)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.